Experiment #38

Transistors (1)
V-I Characteristics

<u>Principles</u>

Introduction

A transistor is categorized as an active circuit element. An active circuit element is one that needs to be energized to become operational. The three (and the only three) passive circuit elements (resistors, capacitors and inductors) have no such needs. The number of active circuit elements is very large. A transistor is not only the most basic but to-date the most important of all active circuit elements.

Unlike passive circuit elements, active circuit elements have three or more terminals. Two of these, of course, are the traditional “in” and “out” terminals. The additional terminal serves to control the flow of electricity by narrowing and widening the path along which electrons travel. The necessary controlling instructions (called “signals”) are fed to this terminal by an external source. In a transistor a copious flow of electrons can easily be controlled by a tiny signal coming from the external source. The output current of the transistor is, therefore, not only an exact replica of the controlling signal but is also much stronger in amplitude.

Function

The main function of a transistor is to faithfully reproduce a weak signal as a strong signal. This is “amplification”. A transistor is typically used as an “amplifier”. Transistors can amplify signals in many different ways.

The Anatomy of a Transistor

(1) Semiconductors

Semiconductor elements differ from all other elements of the periodic table in that they have two unusual criteria for the distribution of their electrons in orbits around their nuclei. Of the known elements, listed in a periodic table, only three elements satisfy these criteria. The elements are (i) carbon, (ii) silicon, and (iii) germanium. The criteria are: (i) there should be exactly four electrons in the valence orbit, and (ii) all other (inner) orbits should be filled to capacity. The capacity of an orbit to hold electrons is given by the formula: $2n^2$ where n is the serial number of the orbit beginning with the one closest to the nucleus. In general, atoms can have any number of electrons in the valence orbit up to a maximum of eight electrons. Again it is a common practice for atoms to begin placing electrons in higher orbits before the lower ones are filled to capacity. With such wide latitude given to the atoms, it is really surprising to find that three of them have chosen such strict distribution pattern for their electrons.

The electron-distribution of the three semiconductor elements is shown in Table (1)

Table 1 Electron-Distribution of Semiconductor Elements

#	name	symbol	Total number of electrons	maximum # of electrons in the n^{th} orbit $2n^2$			valence orbit
				$n = 1$ 2 electrons	$n = 2$ 8 electrons	$n = 3$ 18 electrons	
1	carbon, C	$_{12}C^{6}$	6	2	N/A	N/A	4
2	silicon, Si	$_{28}Si^{14}$	14	2	8	N/A	4
3	germanium, Ge	$_{73}Ge^{32}$	32	2	8	18	4

(2) Stable Chemical Configuration

It is common knowledge that for a stable chemical configuration, atoms need to have eight electrons in their valence orbits, either all of their own or shared with other atoms. Elements that have eight of their own electrons in their valence orbits are known as the “noble” elements. Being in stable chemical configuration, noble elements hardly ever react chemically with other elements and are, therefore, also known as “inert” elements (mostly gases). Helium and neon may be cited as two examples. Those elements that have less than four electrons in their valence orbits, tend to give them away to other suitable elements, so that together they acquire the stable chemical configuration. These are “electro-positive” elements. Sodium is a good example. Those elements, on the other hand, that have more than four electrons in their valence orbits, tend to gain more electrons from other suitable elements so that together they acquire the stable chemical configuration. These are “electro-negative” elements. Chlorine is a good example. This leaves us with those elements that have exactly four electrons in their valence orbits. It is logical to expect that they have equal tendency of gaining four electrons or losing them, in order to have a stable chemical configuration.

(3) Covalent Bonds

One way of making stable chemical configuration for these elements was to form “covalent” chemical bonds with one another. One atom shares one each of its four electrons with four neighboring atoms. Carefully grown crystals of semiconductor elements are found to have such structures. This is shown in Fig (1).

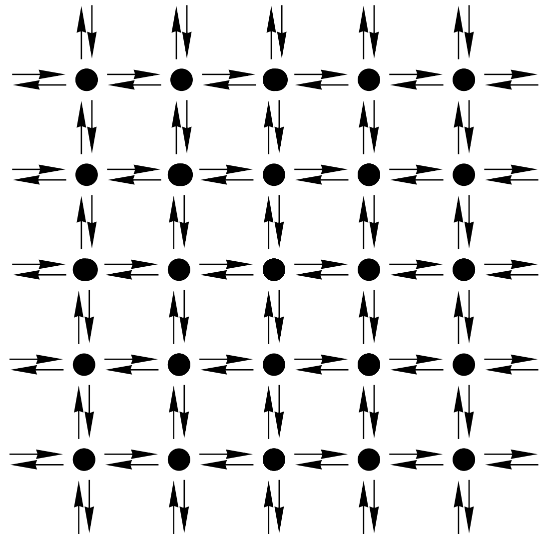


Fig (1) The Covalent Bonds

(4) Free Electrons

Covalent bonds are not very strong. Thermal energy present in the region causes a number of bonds to break, thereby producing free electrons. If thermal energy increases (i.e. temperature rises), the number of free electrons also increases. These free electrons behave like the free electrons in metals: (i) they move around randomly and have negligible displacements over sizeable periods of time, (ii) they conduct electricity in the presence of suitable electric fields. There is, however, one significant difference between the semiconductors and the metals. In metals conduction of electricity decreases as temperature increases; but for semiconductors, conduction improves with the rise in temperature. It is this phenomenon that causes them to have negative temperature coefficients.

This is not something you should pass by casually, even though nothing exciting has happened yet. The most we benefited from this situation (so far) was to develop temperature compensated or temperature independent resistors. With the advent of technology, semiconductors could also be used as sensitive temperature measuring devices.

(5) The Doped Semiconductors: (1) The N-type Semiconductor

A truly revolutionary idea surfaced in some genius's mind. Why not *create* electron-proficient and electron-deficient semiconductor crystal? Electron proficiency could be created by introducing (as an impurity) minute quantities of an element with five electrons in the valence orbit. An impurity atom will replace one of the semiconductor atoms in the crystal. Four electrons of the impurity atom will form four co-valent bonds but the fifth electron of the impurity atom will be left without any attachment. Many of these may be set free by the ever-present thermal energy, thereby creating an abundance of free electrons. These free electrons will help conduct electricity when an electric field is established in the crystal. We may say that electricity will be conducted by negatively charged particles (the electrons) or that the crystal has negative "charge carriers". Examples of elements that may be used as impurity for establishing electron proficiency are Arsenic and Bismuth.

The process of adding impurity is called "doping" and the electron-proficient semiconductor crystal is called an "N-type" semiconductor. Here "N" implies the negativity of the charge carriers.

(6) The Doped Semiconductors: (2) The P-type Semiconductor

An electron-deficient crystal may be created by introducing minute quantities of an element (as an impurity) with three electrons in the valence orbit. An impurity atom will replace one of the semiconductor atoms in the crystal. The three electrons of the impurity atom will form covalent bonds with three neighbors. This will leave the fourth neighbor hungry for another electron to form the much desired covalent bond. Examples of elements that may be used for this purpose are Gallium and Indium.

This electron-deficiency is officially called a "hole". A hole is defined as "absence of an electron" or a "vacancy" and is assigned a positive sign. In reality a hole is a perfect void. The "neighborhood" has imparted it the color "positive". These vacancies or holes are exact counterparts of free electrons in an electron-proficient crystal. Just as free electrons are mobile, the holes are mobile too. As we know, thermal energy keeps breaking a limited number of bonds. When a bond breaks, a free electron is made available but at the same time a new hole is created. These free electrons go into random motion and while moving randomly, they pass by a hole, only to be sucked into it. The hole is filled and a covalent bond is established. We say that the hole moved to the location where the electron came from. This is the mobility of the hole.

Holes keep moving from place to place randomly. This randomness of the movement of holes is a direct consequence of the randomness in which thermal energy keeps breaking bonds and the randomness in which free electrons keep getting re-located. It is interesting to note that holes in an electron-deficient semiconductor are as mobile as electrons in an electron-proficient semiconductor. The electron-deficient semiconductor crystals are found to conducts electricity as efficiently as the electron-proficient semiconductor crystals. The electron deficient semiconductor has positive charge careers.

The electron deficient semiconductor crystal is called a "P-type" semiconductor.

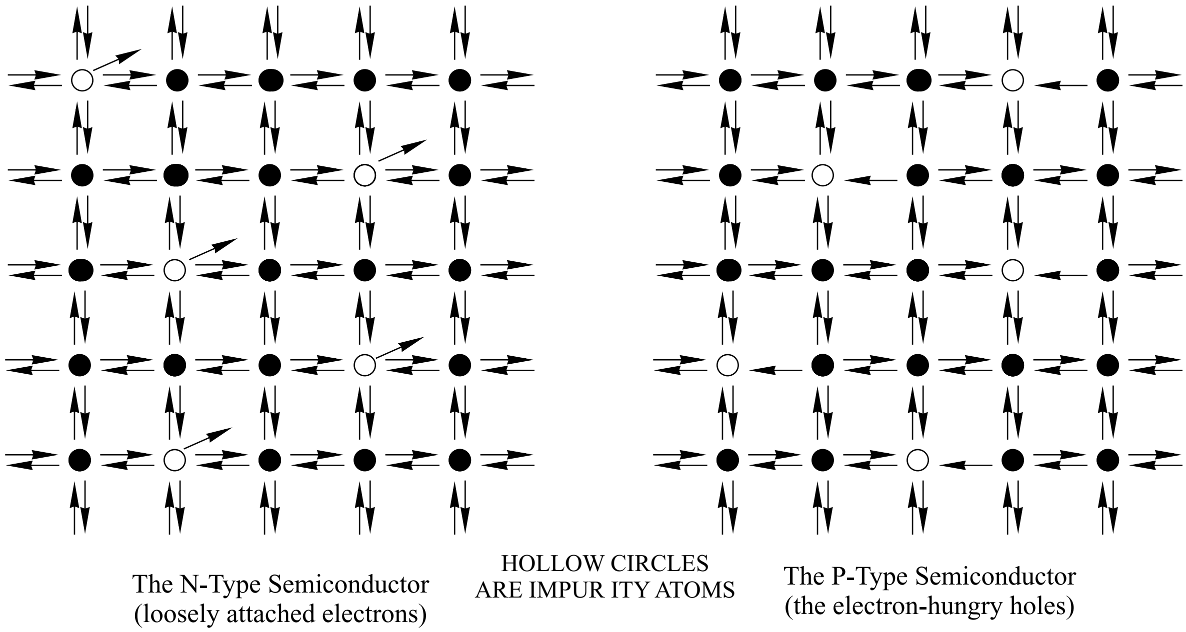


Fig (2) The Electron Proficient and The Electron Deficient Semiconductor Crystals

(7) The P-N Junction

One may be afraid of bringing an N-type and a P-type semiconductor close together thinking that all the free and mobile electrons will rush to fill up the holes and soon enough we shall have an inert piece of a semiconductor crystal that will neither make a temperature compensated resistor nor will it measure temperature. This doesn't happen because free and mobile electrons and holes do not have enough energy to travel long distances.

When we join together a P-type and an N-type semiconductor, all electrons near the boundary (called "junction") rush across the boundary and neutralize as many holes as possible thereby creating a buffer zone where there are neither electrons nor holes. Buffer zone is like the border between two countries, sort of no-man's land. Electrons beyond the buffer zone in the N-type region simply do not have enough energy to cross the (Sahara Desert) buffer zone; so they settle quietly in their normal routine of random motion. We end up having a crystal with abundance of electrons in one part and abundance of holes in the other.

.

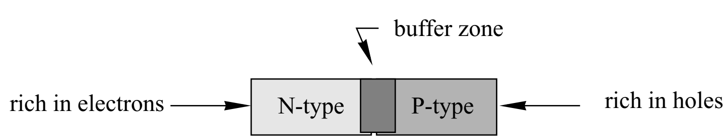


Fig (3) The P-N Junction

(8) P-N Junctions & DC Electricity

We can connect a DC supply to the P-N junction in two different ways. These are shown in Fig (4). In one case, called “forward biasing”, the positive terminal of the battery is connected to the P side of the junction and the negative terminal of the battery is connected to the N side of the junction. In the other case the battery is connected to the junction in an opposite manner. This mode of battery connection is called “reverse biasing”.

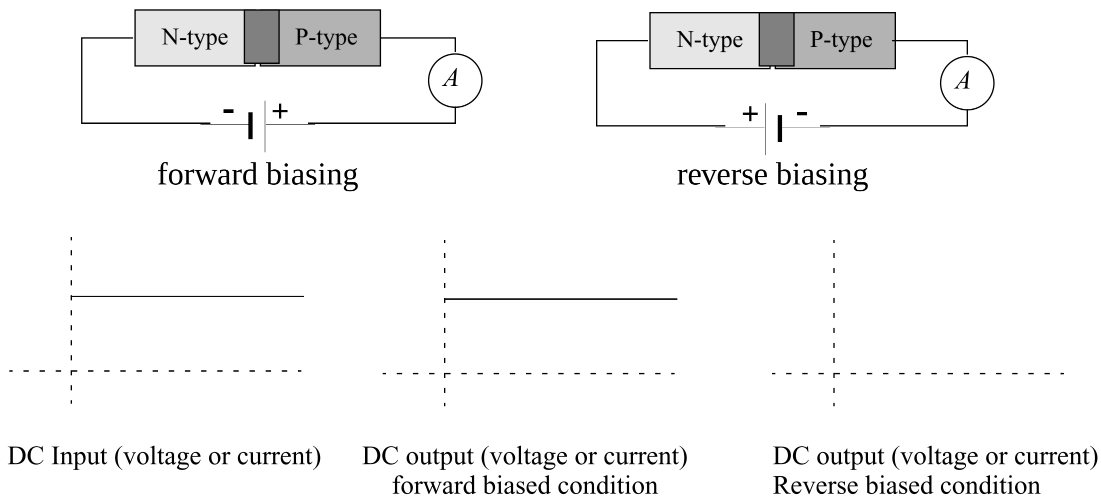


Fig (4) The Biasing of a P-N Junction

In case of forward biasing, the battery sends “combat ready” electrons into the N-type region which is infested with “lazy” electrons that are roaming around aimlessly. These electrons are pushed by the repulsive force of the incoming electrons, away from the incoming electrons, toward the P-N junction and beyond. They move into the P-type territory and fill up the holes thereby re-establishing the covalent bonds. At the same time, the strong positive potential supplied by the battery at the P-end, causes a whole lot of covalent bonds to break up and release electrons. The positive potential (acting like a vacuum pump) sucks out these electrons thereby creating more room for the electrons coming from the N-region of the junction. This results in a continuous flow of electrons from N-type region of the crystal, across the junction, into the P-type region. An ammeter placed in the circuit shows a distinct current flow.

In case of reverse bias, the battery sends the “combat ready” electrons into the P-type territory where they just fill up the holes and establish covalent bonds. At the same time the positive potential at the N-type region of the junction causes the lazy electrons to be attracted toward the positive potential, away from the junction. The result is that there is no traffic across the junction. The circuit is not completed and an ammeter placed in the circuit will indicate **zero** current.

(9) P-N Junctions & AC Electricity

In AC electricity, the source keeps switching the direction of flow of electrons. In one half of a cycle, electrons are sent out from the left terminal of the source and they return to the source through its right terminal. In the other half of the cycle, electrons are sent out from the right terminal of the source and they return to the source through its left terminal. When electrons are sent out from the left side terminal of the source, the P-N junction gets forward biased and a sinusoidal current (positive half only) flows in the circuit. When electrons are sent out from the right side terminal of the source, the P-N junction gets reversed biased and no current flows in the circuit. We, therefore, do not get a continuous flow of current. The junction will transmit only the positive halves of the Ac electricity!

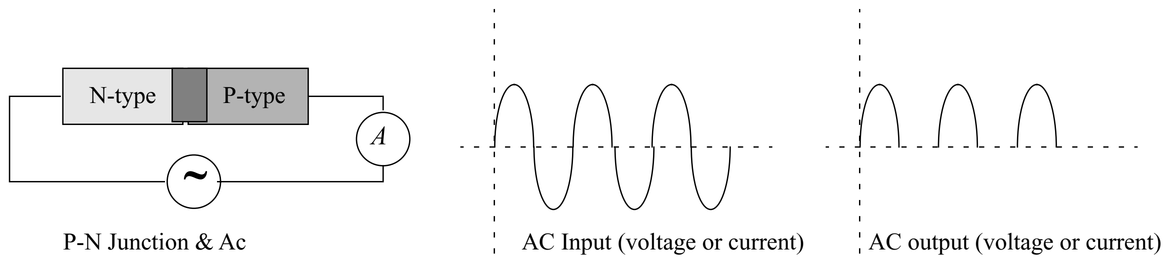


Fig (5) P-N Junction & AC Electricity

(10) P-N Junctions As Diodes

The P-N junction conducts electricity only when it gets forward biased. The junction allows current to flow through it in one direction only. Thus it serves the same purpose as a vacuum tube diode valve. It is, therefore, extensively used as a diode and is termed "Solid State Diode". It has only two terminals; one for electricity to come in and the other for it to leave. The electrical symbol of the diode is shown below, in Fig (6).

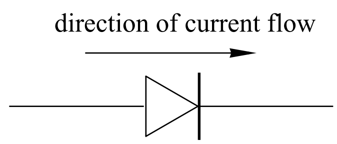


Fig (6) The Solid State Diode

(11) The Transistor

The next step is to consider combining together three crystals of P and N types. We have two possibilities: sandwich a P-type crystal in between two N-type crystals or sandwich an N-type crystal in between two P-type crystals. These possibilities are shown in Fig (7).

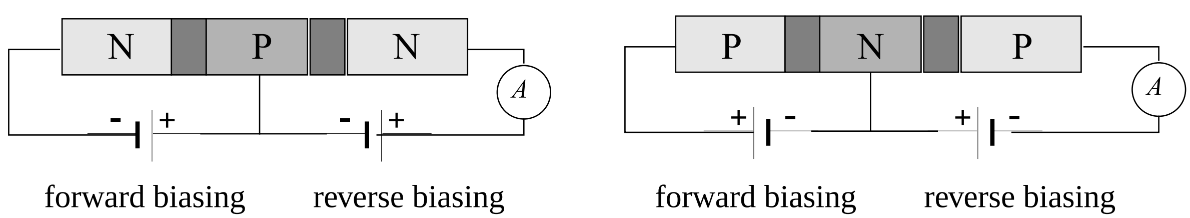


Fig (7) Transistor Biasing

Consider the NPN type transistor. The left hand side junction (the NP junction) being forward biased will send electrons into the P-region. More electrons are injected from the battery in the P-region and the positive potential at extreme right pull all electrons out of the transistor. An ammeter, placed in the circuit, will show a steady flow of current.

A controlling signal when applied at the central P-region, will attract more electrons if it were to be positive and thereby will enhance the electron flow. It will deter electrons from coming in, if it was negative, thereby restricting the electron flow.

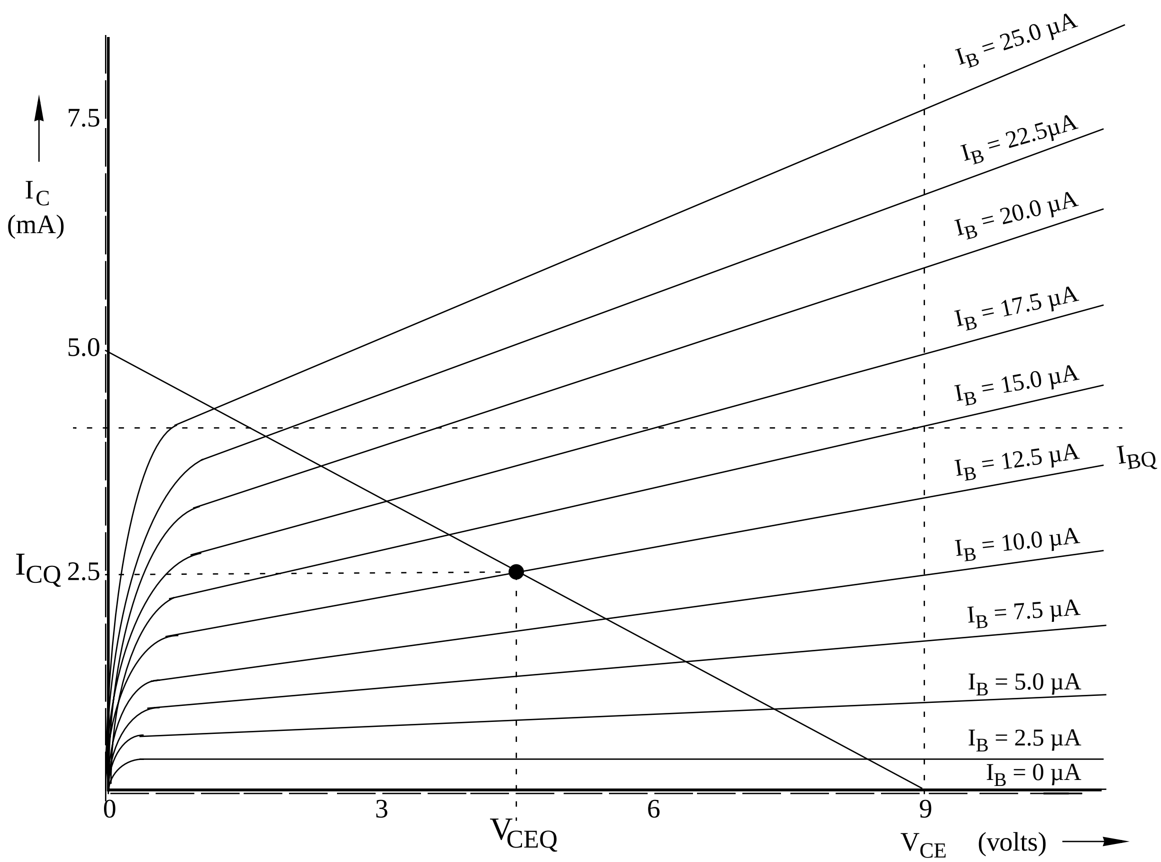


Fig (8) The Transistor Characteristics and the Load Line

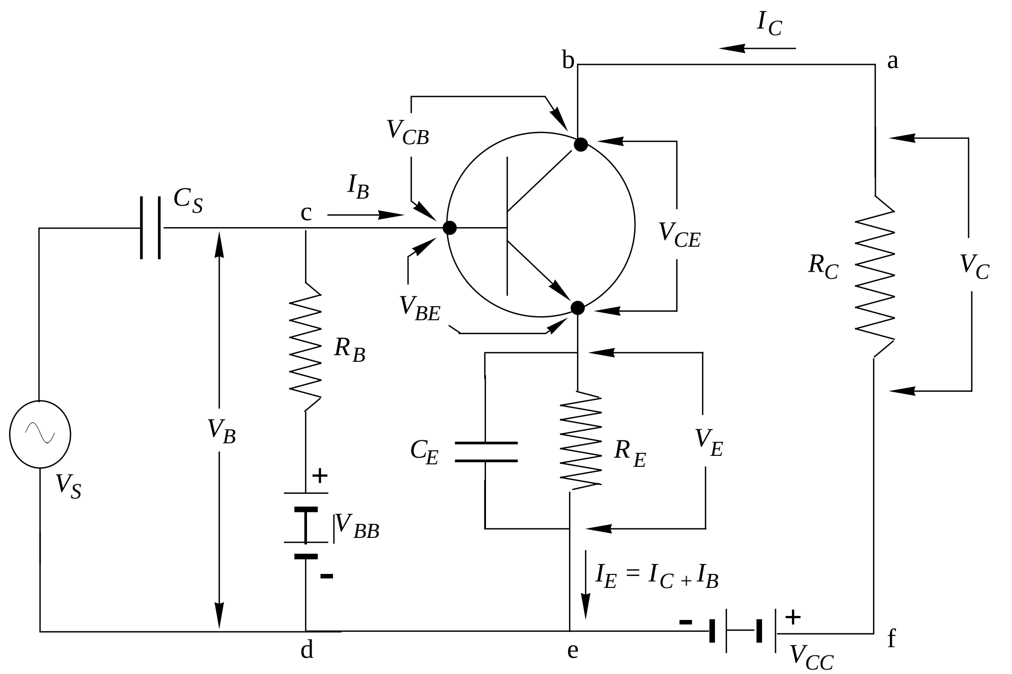


Fig (9) An NPN Transistor Amplifier

Having drawn the set of characteristic curves, draw a horizontal line parallel to the x-axis from the knee of the uppermost curve. Next draw a vertical line parallel to the y-axis from the voltage that you wish to use as V_{CC}. One then selects an "operating point" for the transistor, somewhere in the middle of the rectangle. The values of I_C, I_B and V_{CE} at this point are called I_{CQ}, I_{BQ} and V_{CEQ}. While choosing the operating point one (naturally) looks for convenient values of I_{CQ}, I_{BQ} and V_{CEQ}. The operating point is usually given the name "Q-point" where "Q" stands for "quiescent". If the amplifier is correctly designed then, in the absence of an external signal, the values of I_C, I_B and V_{CE} (called I_{CQ}, I_{BQ} and V_{CEQ}) in the circuit will match the ones at the Q-point, on our graph. The quiescent point is the actual point around which the transistor operates. All input and output signals must be contained within the rectangle, if signal distortions are to be avoided.

The ratio I_{CQ} / I_{BQ} at the Q-point is called β or h_{fe}. This tells us the factor by which the base current is amplified by the transistor. Thus β or h_{fe} is the "current amplification factor" of the transistor and is an important commodity. It also tells us how small the base current is, as compared to the collector current. In fact, due to the smallness of I_B, one can approximate $I_E = I_C + I_B$ to $I_E = I_C$.

Selection of the operating point, in the manner described above, immediately gives us the values of V_{CC}, I_{CQ}, I_{BQ} and V_{CEQ}. One then proceeds to find R_C and R_E. To do so, we apply Kirchhoff's loop rule to the loop "abefa", (external signals being absent); we get:

$$\begin{aligned} V_{CC} &= V_C + V_{CEQ} + V_E \\ &= V_{CEQ} + I_{CQ}R_C + I_{EQ}R_E \end{aligned}$$

Since I_E can be approximated as I_C, we may write the above as:

$$V_{CC} = V_{CEQ} + (R_C + R_E)I_C$$

Now lump together R_C and R_E. Call it R_L

$$V_{CC} = V_{CEQ} + R_L I_{CQ} \qquad \text{..........(1)}$$

This equation can be solved for R_L:

$$R_L = \frac{V_{CC} - V_{CEQ}}{I_{CQ}}$$

The same purpose is served by simply drawing a line from the selected position of V_{CC} to the selected operating point and extending it to meet the y-axis. Why should the slope of R_L be negative? Let's re-arrange Eqn (1):

$$V_{CEQ} = V_{CC} - R_L I_C \qquad \text{..........(2)}$$

It is easily noted that this equation is of the form of the equation of a battery: $V_{AB} = \varepsilon - rI$.

Having drawn the load line and hence determined the value of R_L, one needs to split it between R_C and R_E. To do this we use a rule of thumb which states that R_E should be roughly one-tenth of R_L. This sets R_C as nine-tenth of R_L.

In our diagram of the characteristics of an NPN transistor (to serve solely as an example), we find the following values of I_C, I_B and V_{CE} at the selected operating point:

$$V_{CC} = 9.0 \text{ V}$$

$$I_{CQ} = 2.5 \text{ mA}$$

$$I_{BQ} = 12.5 \text{ A}$$

$$V_{CEQ} = 4.5 \text{ V}$$

$$h_{fe} = \frac{2.5\times10^{-3}}{12.5\times10^{-6}} = 200$$

$$R_L = \frac{(9-0)}{(5\times10^{-3}-0)} = 1800 \text{ ‰}$$

$$R_E = \frac{1800}{10} = 180 \text{ ‰}$$

$$R_C = (1800 - 180) = 1620 \text{ ‰}$$

Next step to find R_B and V_{BB}. Here we use another rule of thumb. It suggests that R_B should approximately be fifteen times larger than R_E. We require R_B to be large so that the incoming signal does not get drained through R_B. If the value of R_B, as found with the rule of thumb isn't large enough, we may use our discretion and increasing it. Once R_B has been selected, we may proceed to determine V_{BB}. Applying Kirchhoff's loop rule to the loop "cdec" we get:

$$V_{BB} = V_{BE} + R_B I_B + R_E I_E \qquad \text{..........(3)}$$

Now $I_E = I_B + I_C$ where $I_B = \frac{I_C}{h_{fe}}$. Inserting these values in Eqn (3) we get:

$$V_{BB} = V_{BE} + R_B\left(\frac{I_C}{h_{fe}}\right) + R_E\left(I_C + \frac{I_C}{h_{fe}}\right)$$

$$= V_{BE} + \left(\frac{R_B}{h_{fe}}\right)I_C + R_E\left(\frac{I_C \times h_{fe} + I_C}{h_{fe}}\right)$$

$$= V_{BE} + \left(\frac{R_B}{h_{fe}}\right)I_C + R_E\left(\frac{h_{fe} + 1}{h_{fe}}\right)I_C$$

Letting $(h_{fe} + 1) \approx h_{fe}$ (since h_{fe} is quite large as compared to 1), we find that $\left(\frac{h_{fe} + 1}{h_{fe}}\right) = 1$. We get:

$$V_{BB} = V_{BE} + \left(\frac{R_B}{h_{fe}}\right)I_C + R_E I_C$$

Finally, we find an expression to determine V_{BB}:

$$V_{BB} = V_{BE} + \left(\frac{R_B}{h_{fe}} + R_E\right)I_C \qquad \text{..........(4)}$$

We use the above to calculate R_B and V_{BB} for our amplifier. The value of V_{BE} depends on the type of semiconductor material used. It has a value of $0.7\ \text{V}$ for silicon and $0.2\ \text{V}$ for germanium. We shall assume that our transistor is made of silicon. We find:

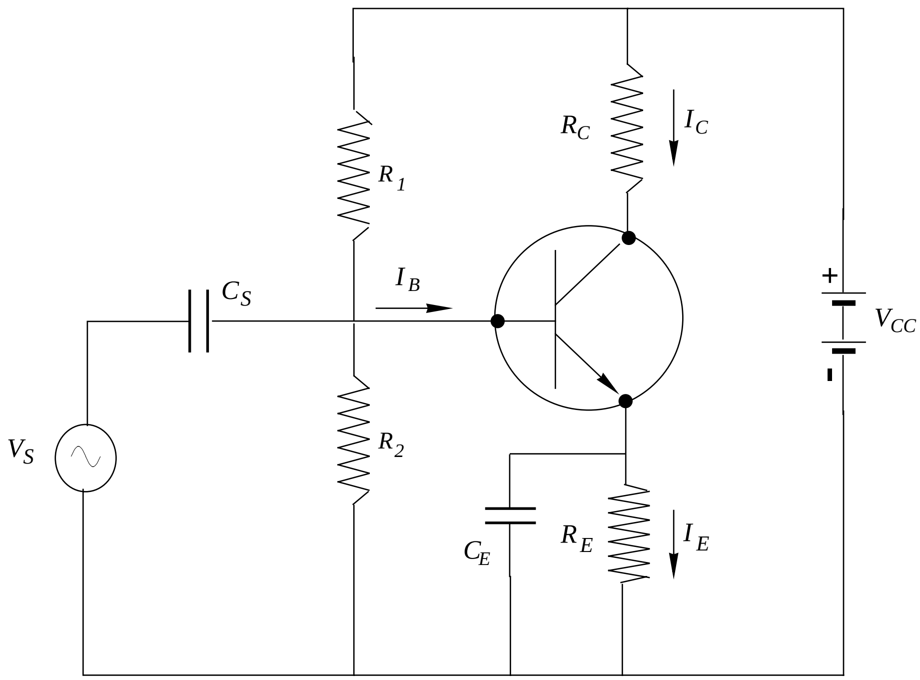
$R_B = 15 \times 180 = 2700$ ‰

This value of R_B is not large enough so let's double it. Let

$R_B = 5400$ ‰

$V_{BB} = 0.7 + \left(\frac{5400}{200} + 180\right) \times 2.5\times10^{-3} = 1.22\ \text{V}$

As we plan to use only one battery, we ned to obtain V_{BB} from V_{CC} by using a potential divider, consisting of two resistors R_1 and R_2, as shown in Fig (10).



The necessary formulae are:

$$V_{BB} = \left(\frac{R_2}{R_1 + R_2}\right) V_{CC} \qquad \text{..........(5)}$$

and

$$R_B = \frac{R_1 R_2}{R_1 + R_2} \qquad \text{..........(6)}$$

We solve these two equations to find R_1 and R_2. From Eqn (6) we find:

$$\frac{R_B}{R_1} = \frac{R_2}{R_1 + R_2}$$

Inserting this value of $\frac{R_2}{R_1 + R_2}$ in Eqn (5), we get:

$$V_{BB} = \left(\frac{R_B}{R_1}\right) V_{CC}$$

Or:

$$R_1 = R_B\left(\frac{V_{CC}}{V_{BB}}\right) \qquad \text{..........(7)}$$

From Eqn (6) cross multiplying and solving for R_2, we get:

$$R_2 = \frac{R_B R_1}{R_1 - R_B} \qquad \text{..........(8)}$$

We now find values of R_1 and R_2 for our example:

$$R_1 = 5400 \times \frac{9}{1.22} = 39836 \text{ ‰} \approx 40 \text{ K‰}$$

$$R_2 = \left(\frac{5400 \times 39836}{39836 - 5400} = 6246.8 \text{ ‰}\right) \approx 6.3 \text{ K‰}$$

Next we calculate C_E. The purpose of this capacitor is to bye-pass the AC (incoming) signal reaching the emitter-base junction so that the operating point is not disturbed. The reactance of this capacitor for the lowest frequency of interest should a tenth of the emitter resistor R_E.

$$\chi_C = \frac{1}{2\pi f_o C_E}$$

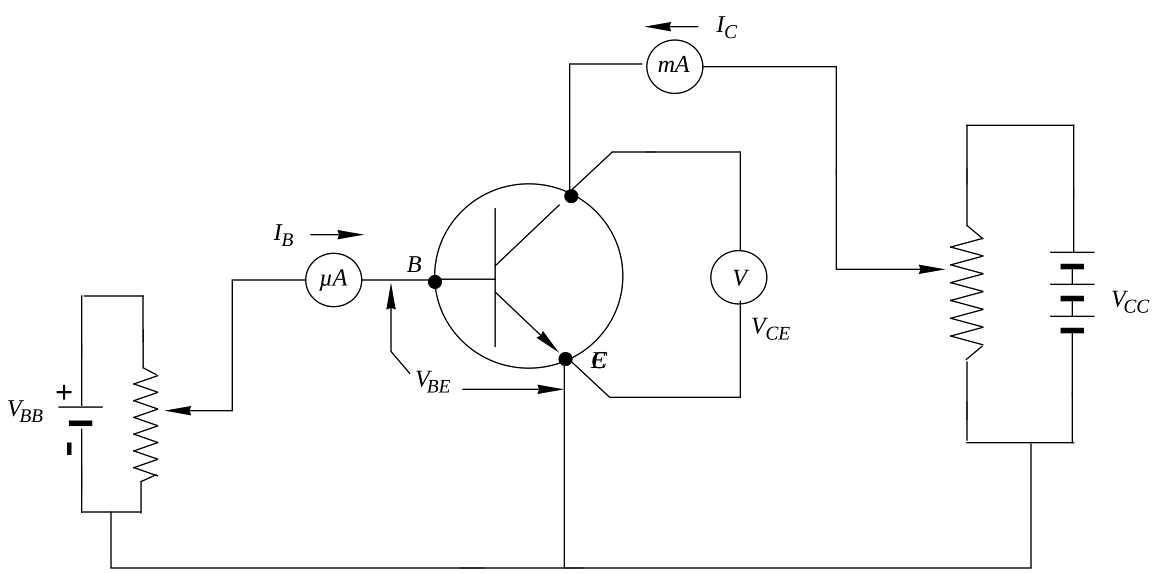
Also

$$\chi_C = 0.1 \times R_E$$

Letting $f_o = 50 \text{ Hz}$, we solve for C_E:

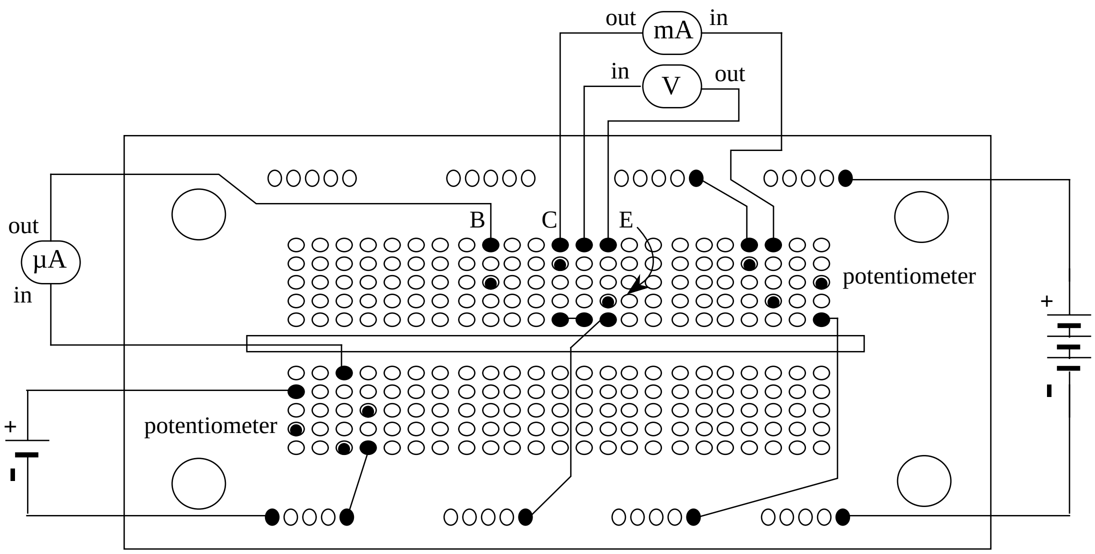
$$0.1 \times 180 = \frac{1}{2\pi \times 50 \times C_E}$$

$$C_E = 177 \text{ F}$$



Procedure

1. Assemble the circuit on a bread board. A sketch is given below:



Circuit Diagram for Obtaining Data for the V-I Chatacteristic Curves of a Transistor

2. Please note that the right hand side potentiometer is for setting values of V_{CE} while the one on the left hand side is for controlling V_{BE}.
3. You should clearly mark the two ammeters. The one for measuring I_C should be on the right hand side of the transistor while that for measuring I_B should be placed on the left side. The only voltmeter used is for measuring V_{CE}. Polarities of all meters are shown. The “in” terminals are usually red while the “out” terminals are black.